



# UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE  
United States Patent and Trademark Office  
Address: COMMISSIONER FOR PATENTS  
P.O. Box 1450  
Alexandria, Virginia 22313-1450  
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
09/998,959	11/01/2001	Bishnu Saroop Atal	2000-0606	6485

7590 04/08/2005

SAMUEL H. DWORETSKY  
ATT&T CORP  
ROOM 2A-207  
ONE ATT&T WAY  
BEDMINSTER, NJ 07921

EXAMINER

ALBERTALLI, BRIAN LOUIS

ART UNIT	PAPER NUMBER
----------	--------------

2655

DATE MAILED: 04/08/2005

Please find below and/or attached an Office communication concerning this application or proceeding.

<b>Office Action Summary</b>	<b>Application No.</b>	<b>Applicant(s)</b>	
	09/998,959	ATAL, BISHNU SAROOP	
	<b>Examiner</b>	<b>Art Unit</b>	
	Brian L. Albertalli	2655	

**-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --**

**Period for Reply**

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If the period for reply specified above is less than thirty (30) days, a reply within the statutory minimum of thirty (30) days will be considered timely.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

**Status**

- 1) ☒ Responsive to communication(s) filed on 26 November 2004.
- 2a) ☐ This action is **FINAL**.                      2b) ☒ This action is non-final.
- 3) ☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

**Disposition of Claims**

- 4) ☒ Claim(s) 1-36 is/are pending in the application.
- 4a) Of the above claim(s) \_\_\_\_\_ is/are withdrawn from consideration.
- 5) ☐ Claim(s) \_\_\_\_\_ is/are allowed.
- 6) ☒ Claim(s) 1-36 is/are rejected.
- 7) ☐ Claim(s) \_\_\_\_\_ is/are objected to.
- 8) ☐ Claim(s) \_\_\_\_\_ are subject to restriction and/or election requirement.

**Application Papers**

- 9) ☐ The specification is objected to by the Examiner.
- 10) ☒ The drawing(s) filed on 01 November 2001 is/are: a) ☐ accepted or b) ☒ objected to by the Examiner.  
Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).  
Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).
- 11) ☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

**Priority under 35 U.S.C. § 119**

- 12) ☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).
- a) ☐ All    b) ☐ Some \*    c) ☐ None of:
1. ☐ Certified copies of the priority documents have been received.
  2. ☐ Certified copies of the priority documents have been received in Application No. \_\_\_\_\_.
  3. ☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

\* See the attached detailed Office action for a list of the certified copies not received.

**Attachment(s)**

- |  |   |
|--|---|
| 1) <input checked="" type="checkbox"/> Notice of References Cited (PTO-892)  | 4) <input type="checkbox"/> Interview Summary (PTO-413)<br>Paper No(s)/Mail Date. _____ |
| 2) <input type="checkbox"/> Notice of Draftsperson's Patent Drawing Review (PTO-948)                                   | 5) <input type="checkbox"/> Notice of Informal Patent Application (PTO-152)             |
| 3) <input type="checkbox"/> Information Disclosure Statement(s) (PTO-1449 or PTO/SB/08)<br>Paper No(s)/Mail Date _____ | 6) <input type="checkbox"/> Other: _____  |

## **DETAILED ACTION**

### ***Response to Arguments***

1. Applicant's arguments with respect to claims 1-36 have been considered but are moot in view of the new ground(s) of rejection.

### ***Drawings***

2. The objection to the drawings in the previous office action was incorrect. This objection to the drawings is withdrawn.

However, the drawings are still objected to, for containing a reference number in the specification that is not in the drawings. On page 14, paragraph 52, line 2, the specification lists reference number 108 as appearing in Fig. 7a, but this reference is not in Fig. 7a. Similarly, although not specifically mentioned in the specification, it appears that the /ay/ phoneme in Fig. 7b and Fig. 7c should be labeled with reference number 108 as well.

The objection to the drawings will not be held in abeyance.

### ***Claim Rejections - 35 USC § 102***

3. The following is a quotation of the appropriate paragraphs of 35 U.S.C. 102 that form the basis for the rejections under this section made in this Office action:

A person shall be entitled to a patent unless –

(b) the invention was patented or described in a printed publication in this or a foreign country or in public use or on sale in this country, more than one year prior to the date of application for patent in the United States.

Art Unit: 2655

4. Claims 25, 30, and <sup>35</sup>31 are rejected under 35 U.S.C. 102(b) as being anticipated by Hattori (U.S. Patent 5,140,668).

In regard to claims 25 and 35, Hattori discloses a method and medium storing a program for recognizing speech using a database of stored phonemes converted into n-dimensional space (see Title), the method comprising:

receiving a received phoneme (input phoneme pattern, column 3, lines 11-14);

converting the received phoneme to n-dimensional space (see Fig. 3, input phonemes 14 and 15 are converted to feature space);

comparing the received phoneme to each of the stored phonemes in n-dimensional space (the input vectors are compared to all of the reference vectors, column 3, lines 20-25); and

recognizing the received phoneme according the comparison of the received phoneme to each of the stored phonemes (the comparison between the input vectors and reference vectors is a distance measure, column 3, lines 20-25 and equations 1 and 2; the smallest distance measure produces the recognition result, column 3, lines 2-8).

In regard to claim 30, Hattori discloses a system for recognizing phonemes (Fig. 4), the system using a database of stored phonemes for comparison with received phonemes (reference phoneme pattern memory 32, column 3, lines 14-16), the stored phonemes having been converted into n-dimensional space (see Fig. 3, stored patterns A, B, and C), the system comprising:

a recording element that receives a phoneme (input terminal 30, column 3, lines 11-14);

a computer that converts the received phoneme into n-dimensional space (see Fig. 3, input phonemes 14 and 15 are converted to feature space), wherein the computer compares in the n-dimensional space the received phoneme with each phoneme in the database of stored phonemes (distance calculator 33 compares the input vectors to all of the reference vectors, column 3, lines 20-25).

In regard to claim 31, Hattori discloses the computer recognizes the received phoneme using the comparison in the n-dimensional space of the received phoneme with each phoneme in the database of stored phonemes (the comparison between the input vectors and reference vectors is a distance measure, column 3, lines 20-25 and equations 1 and 2; the smallest distance measure produces the recognition result, column 3, lines 2-8).

### ***Claim Rejections - 35 USC § 103***

5. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

(a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negated by the manner in which the invention was made.

6. Claims 1, 2, and 5 are rejected under 35 U.S.C. 103(a) as being unpatentable over Casey (U.S. Patent 6,321,200), in view of Lilly et al. (*Robust Speech Recognition*

*Using Singular Value Decomposition Based Speech Enhancement*), and further in view of Smith et al. (*Template Adaptation in a Hypersphere Word Classifier*).

In regard to claims 1 and 36, Casey discloses a method and a medium storing a program for extracting features from a set of phonemes (the method disclosed by Casey is used for pattern recognition techniques in the domain of speech phonemes, column 4, lines 25-29), comprising:

(1) determining a phoneme vector as a time-frequency representation of the class phoneme;

(2) dividing the phoneme vector into phoneme segments;

(3) assigning each phoneme segment into a plurality of phoneme parameters;

and

(4) expanding each phoneme segment and plurality of phoneme parameters into an expanded stored-phoneme vector with expanded vector parameters

As described in the specification, the steps of “determining a phoneme vector”, “dividing the phoneme vector”, “assigning each phoneme segment into a plurality of phoneme parameters”, and “expanding each phoneme segment” as claimed in both the training and recognizing processes, is nothing more than dividing an input phoneme into 25 msec sections *represented using 32 mel-spaced filters*, and converging (expanding) these vectors into a high-dimensional “expanded stored phoneme vector with expanded vector parameters” (page 15, paragraph 53). The expanding step is a simple concatenation of the 5 segments of phoneme parameters (5 segments\*32 filters = 160 dimensions).

Casey discloses a method that represents the input signal as a time-frequency representation (forty to fifty filtered signals are output at Fig. 1, step 110, column 3, lines 1-10);

dividing the input signal into segments (20 msec segments, step 120, column 3, lines 11-13); and

assigning each segment a plurality of parameters (each time instance N is assigned the 40-50 parameters output from the filters, see observation matrix 121); and

expanding each segment into an expanded stored matrix with expanded vector parameters (each of the columns of the observation matrix 121 represents an expanded vector and these are concatenated to create the observation matrix as a whole).

While Casey does not specifically call the extracted features “phoneme” features (phoneme vector, phoneme segments, etc.), as discussed above, Casey teaches using the extraction method in the domain of speech phonemes. Therefore, when extracting features from phonemes for use in a speech pattern classification system, the resulting features would clearly be “phoneme features”

Casey further discloses transforming the stored phoneme vector (observation matrix 121) into an orthogonal form using singular value decomposition (step 130, column 3, lines 16-31).

Casey then takes further steps (namely, the ICA analysis in step 140) to produce a signal used for classification.

Casey does not disclose using the results of the singular value decomposition as parameters for a pattern classifier.

Lilly et al. disclose using singular value decomposition (SVD) on extracted features from speech reduces the noise levels in the speech signal (which is common in real world speech recognition applications) and improves the recognition performance (page 259-260, section 4.2).

It would have been obvious to one of ordinary skill in the art at the time of invention to modify Casey to use the output of the SVD (reduced dimensionality matrices 131) as the extracted features used for phoneme recognition, since using SVD in a front-end preprocessor (feature extractor) before recognition improves the recognition performance (page 260, section 5, lines 1-4).

Neither Casey nor Lilly et al. disclose that the extracted features are used to train class phonemes or recognize class phonemes, wherein the recognition was determined by:

determining a first distance associated with the orthogonal form of the expanded received-signal vector and a second distance associated respectively with each orthogonal form of the expanded stored-phoneme vectors; and  
Recognizing the received phoneme according to a comparison of the first distance with the second distance.

Smith et al. discloses a method of speech recognition that utilizes a hyperspheres as templates for word recognition. Although Smith et al. uses word templates, the difference between using phoneme templates and word templates is simply a matter of how many segments (roughly 20 msec windows) are used as the templates. This fact is discussed in the applicant's specification (page 18, paragraph



59). The method includes a training phase (see page 565, Template Generation section) and a recognition phase (Matching section).

The recognition phase of Smith's method comprises:

Determining a first distance associated with the received-signal vector and a second distance associated respectively with the expanded stored-phoneme vectors (template of the representing the input is compared with each of the templates in the vocabulary); and

Recognizing the received phoneme according to a comparison of the first distance with the second distance (the vocabulary template with the best score is used as the recognition result, page 565, second column, Matching section).

It would have been obvious to one of ordinary skill in the art at the time of invention to further modify the combination of Casey and Lilly et al., so that the method of extracting features, as discussed in the combination of Casey and Lilly et al., was used to create the template patterns used in speech pattern training and recognition, as disclosed by Smith et al., so that a highly accurate representation of speech could be used in the speech recognizer, thereby increasing the chances of correct recognition results.

In regard to claim 2, Casey discloses transforming the stored-phoneme vector (observation matrix 121) into an orthogonal form using singular-value decomposition (step 130, column 3, lines 16-31).

Casey and Lilly et al., as applied to claim 1, above, do not disclose that transforming the expanded received-signal vector into an orthogonal form using singular-value decomposition conforms the stored-phoneme vector and the expanded received-signal vector into a hypersphere having a center and a radius.

Smith et al. discloses that a stored-phoneme vector and a received-signal vector (templates) are n-dimensional hyperspheres (page 565, Description of the Recognizer section), which applies also to their orthogonal forms. A hypersphere must, by definition, have a center and a radius.

It would have been obvious to one of ordinary skill in the art at the time of invention to modify Casey to conform the stored phoneme vector into a hypersphere, since representing stored-phoneme vectors and received-phoneme vectors as hyperspheres simplifies the process of adapting the stored-phoneme vectors (templates) for better recognition results, as taught by Smith et al. (pages 565-566, Adaptive Training section).

In regard to claim 5, Casey discloses the orthogonal form of the expanded stored-phoneme vector and the expanded received-signal vector each have at least approximately 100 dimensions.

Casey discloses that 40 to 50 spectral parameters are included in each observation matrix, which includes hundreds of samples (column 3, lines 4-7, and lines 13-14). This would correspond to at least a 4000 dimensional (100 samples \* 40

Art Unit: 2655

frequency bands) matrix. The stored-phoneme vector would be created in the same manner as the expanded received-signal vector.

7. Claims 3-4, and 6-7 are rejected under 35 U.S.C. 103(a) as being unpatentable over Casey, in view of Lilly et al., in further view of Smith et al., and further in view of Cooper (*The Hypersphere in Pattern Recognition*).

In regard to claim 3, Casey, Lilly et al., and Smith et al. do not disclose that determining a distance comprises comparing a distance from the center of the hypersphere of the orthogonal form of the expanded received-signal vector with a distance from the center of the hypersphere for each orthogonal form of the expanded stored-phoneme vector.

Cooper discloses that when using a hypersphere as a classification boundary, the distance of an unknown vector  $x$  from the hypersphere is determined by calculating the distance of the unknown vector  $x$  from the center of the hypersphere (page 326, section II, lines 1-10, equation 2).

It would have been obvious to one of ordinary skill in the art at the time of invention to modify the combination of Casey, Lilly et al., and Smith et al. so the distance between a stored-signal phoneme vector orthogonal form and a received-signal phoneme vector orthogonal form was measured comparing the center of the stored-signal phoneme vector and each received-signal phoneme vector, because comparing a threshold with the distance between an unknown and a fixed point, such as

the center of the hypersphere, can be an excellent approximation decision boundary, as taught by Cooper (page 325 lines 21-26).

In regard to claim 4, the combination of Casey, Lilly et al., Smith et al. and Cooper, as applied to claim 3, above, discloses in Cooper that the hypersphere can be used for multiple category classification (page 338-339, Multiple Category section).

Neither Casey, Lilly et al., Smith et al., nor Cooper disclose that the m-shortest differences between the center of the hypersphere of the received-signal vector and the center of the hypersphere for each orthogonal form of the expanded stored-phoneme vectors are recognized as most likely to be associated with the received phoneme.

Official notice is taken that it is notoriously well known in the art to consider several of the best choices as a possible recognition result. In a system that used a distance measurement to determine the similarity between an input phoneme and a stored phoneme, this would correspond to the m-shortest differences between the input phoneme and the stored phoneme.

It would have been obvious to one of ordinary skill in the art at the time of invention to modify the combination of Casey, Lilly et al., Smith et al., and Cooper, so the m-shortest differences (which would be the m-best recognition choices) would be recognized as most likely to be associated with the received phoneme, so that if the shortest distance was determined to be an incorrect recognition result during later processing (such as during word recognition or phrase recognition) the next shortest distance result could be used.

In regard to claims 6 and 7, Casey, Lilly et al., and Smith et al. do not disclose removing the mean value of a stored phoneme or a received phoneme vector.

Cooper discloses that for two distributions having the same mean, the hyperspheres corresponding with those distributions will have the same center. Therefore, the comparison between two distributions with the same mean only requires a calculation of the radius of the hypersphere (page 329, section C, lines 1-3 and page 330, lines 1-4).

It would have been obvious to one of ordinary skill in the art at the time of invention to remove the means of the stored phoneme vectors and the received phoneme vectors so the comparison of the received vectors to the stored vectors would only require a calculation of the radius of each.

8. Claims 8-15 are rejected under 35 U.S.C. 103(a) as being unpatentable over Casey, in view of Lilly et al., in view of Smith et al., in view of Cooper, and further in view of Ostendorf (*A Stochastic Segment Model for Phoneme-Based Continuous Speech Recognition*).

In regard to claim 8, Casey discloses recognizing phonemes (column 4, lines 25-29).

Casey, Lilly et al., Smith et al. and Cooper do not disclose that the phoneme vector determined as a time-frequency representation of the class phoneme is a representation of approximately 125 msec.

Ostendorf discloses a phoneme recognition method that determines a time-frequency representation of a class phoneme that is approximately 125 msec (frames of speech are analyzed every 10 msec, page 1864, second column, third paragraph; and the average number of samples per phoneme is about 10, page 1859, second column, second paragraph, lines 5-10. This corresponds to a time-frequency representation of 100 msec.)

It would have been obvious to one of ordinary skill in the art at the time of invention to modify the combination of Casey, Lilly et al., Smith et al., and Cooper to determine a time-frequency representation of a class phoneme that was approximately 125 msec long, since this is a good approximation of the length of a phoneme, therefore all of the information necessary for recognition of that phoneme would be included in the 125 msec window.

In regard to claim 9, Casey discloses the phoneme vector is divided into approximately 25 msec phoneme segments (20 msec segments, column 3, lines 11-13).

In regard to claim 10, Casey discloses each phoneme segment is assigned approximately 32 phoneme parameters (each 20 msec phoneme segment is assigned 40-50 parameters, column 3, lines 4-7).

In regard to claim 11, the combination of Casey, Lilly et al., Smith et al., Cooper, and Ostendorf would produce and expanded-stored phoneme with approximately 160 parameters.

An approximately 125 msec time-frequency representation (100 msec) as disclosed by Ostendorf, as applied to claim 8, above, would be windowed every 20 msec, as disclosed by Casey (column 3, lines 11-13), each window having about 40 spectral parameters, as disclosed by Casey (column 3, lines 4-7). This would result in a vector of 200 parameters.

In regard to claim 12, Casey, Lilly et al., Smith et al. and Cooper do not disclose that the received-signal vector determined as a time-frequency representation of the class phoneme is a representation of approximately 125 msec.

Ostendorf discloses a phoneme recognition method that determines a time-frequency representation of a received-signal vector that is approximately 125 msec (frames of speech are analyzed every 10 msec, page 1864, second column, third paragraph; and the average number of samples per phoneme is about 10, page 1859, second column, second paragraph, lines 5-10. This corresponds to a time-frequency representation of 100 msec.)

It would have been obvious to one of ordinary skill in the art at the time of invention to modify the combination of Casey, Smith et al., and Cooper to determine a time-frequency representation of a received-signal vector that was approximately 125 msec long, since this is a good approximation of the length of a phoneme, therefore all

Art Unit: 2655

of the information necessary for recognition of that phoneme would be included in the 125 msec window.

In regard to claim 13, Casey discloses the received phoneme vector is divided into approximately 25 msec phoneme segments (20 msec segments, column 3, lines 11-13).

In regard to claim 14, Casey discloses each phoneme segment is assigned approximately 32 phoneme parameters (each 20 msec phoneme segment is assigned 40-50 parameters, column 3, lines 4-7).

In regard to claim 15, the combination of Casey, Lilly et al., Smith et al., Cooper, and Ostendorf, as applied to claim 12, above, would produce an expanded received-signal vector with approximately 160 parameters.

An approximately 125 msec time-frequency representation (100 msec) as disclosed by Ostendorf, as applied to claim 8, above, would be windowed every 20 msec, as disclosed by Casey (column 3, lines 11-13), each window having about 40 spectral parameters, as disclosed by Casey (column 3, lines 4-7). This would result in a vector of 200 parameters.

9. Claims 16, 18-22, 26-29, and 32-34 are rejected under 35 U.S.C. 103(a) as being unpatentable over Hattori, in view of Kuhn et al. (U.S. Patent 4,292,471).



In regard to claim 16, Hattori discloses a method of recognizing speech patterns, the method using stored phonemes (stored in reference pattern memory 32), the method comprising:

converting each stored phoneme into n-dimensional space (see Fig. 3, stored patterns A, B, and C);

sampling speech patterns to obtain at least one sampled phoneme (input phoneme pattern, column 3, lines 11-14);

converting each of the at least one sampled phonemes into the n-dimensional space (see Fig. 3, input phonemes 14 and 15 are converted to feature space); and

comparing a distance to the sampled phoneme with a distance to each of the phonemes of the converted plurality of phonemes (distance calculator 33 compares the input vectors to all of the reference vectors, column 3, lines 20-25).

Hattori does not disclose comparing distance from the center of the n-dimensional space.

Kuhn et al. disclose a method for comparing distances between received speech patterns in n-dimensional space (where  $n > 2$ , column 5, lines 8-18), comprising comparing the distance of a speech sample to the center point of a reference pattern, (column 15, lines 8-18).

It would have been obvious to one of ordinary skill in the art at the time of invention to modify Hattori to extract more features from an input phoneme, thereby creating a higher dimensional pattern, and to use a distance measurement to the center of the n-dimensional space, since the center point provides an "average" of the

extracted features to make a comparison to. This would ensure that the recognition of a phoneme would be dependent on the overall pattern of an input phoneme, rather than being adversely affected by a small number of divergent features.

In regard to claim 18, Hattori discloses storing the converted phonemes before sampling speech patterns.

The converted phonemes must necessarily be stored before sampling the speech patterns, otherwise there would be no reference phonemes to be compared to for recognition.

In regard to claim 19, Hattori do not disclose "n" is at least 100.

Kuhn et al. disclose that, in practice, higher dimensional characteristics are derived from speech samples so that a characteristic volume of the speech samples can be created (column 5, lines 14-18).

It would have been obvious to one of ordinary skill in the art at the time of invention to modify Hattori to make "n" at least 100 dimensions, in order to provide a highly accurate characteristic volume for input phoneme samples, thereby increasing recognition accuracy.

In regard to claim 20, Hattori disclose comparing a first distance to a first point associated from the received phoneme with a second distance associated in turn with each of the stored phonemes (the distance between an received vector and each

Art Unit: 2655

reference vector is determined, and the distances are compared to find the minimal distance, which is used as the recognition result, column 3, column 3, lines 2-8 and lines 20-42).

Hattori does not disclose comparing distance from the center of the n-dimensional space.

Kuhn et al. disclose a method for comparing distances between received speech patterns in n-dimensional space (where  $n > 2$ , column 5, lines 8-18), comprising comparing the distance of a speech sample to the center point of a reference pattern, (column 15, lines 8-18).

It would have been obvious to one of ordinary skill in the art at the time of invention to modify Hattori to extract more features from an input phoneme, thereby creating a higher dimensional pattern, and to use a distance measurement to the center of the n-dimensional space, since the center point provides an "average" of the extracted features to make a comparison to. This would ensure that the recognition of a phoneme would be dependent on the overall pattern of an input phoneme, rather than being adversely affected by a small number of divergent features.

In regard to claim 21, Hattori discloses recognizing the sampled phoneme as the stored phoneme associated with the smallest difference between the distance from the n-dimensional space to the sampled phoneme with the distance of the n-dimensional space to each of the converted phonemes (the comparison between the input vectors and reference vectors is a distance measure, column 3, lines 20-25 and equations 1

Art Unit: 2655

and 2; the smallest distance measure produces the recognition result, column 3, lines 2-8).

Hattori does not disclose comparing distance from the center of the n-dimensional space.

Kuhn et al. disclose a method for comparing distances between received speech patterns in n-dimensional space (where  $n > 2$ , column 5, lines 8-18), comprising comparing the distance of a speech sample to the center point of a reference pattern, (column 15, lines 8-18).

It would have been obvious to one of ordinary skill in the art at the time of invention to modify Hattori to extract more features from an input phoneme, thereby creating a higher dimensional pattern, and to use a distance measurement to the center of the n-dimensional space, since the center point provides an "average" of the extracted features to make a comparison to. This would ensure that the recognition of a phoneme would be dependent on the overall pattern of an input phoneme, rather than being adversely affected by a small number of divergent features.

In regard to claim 22, Hattori does not disclose the n-dimensional space is hyperspherical.

Kuhn et al. disclose an n-dimensional space that is hyperspherical (see Fig. 1, stored patterns form a circle with a center point M and a radius D, column 5, lines 42-45; wherein in practice more than 3 dimensional representations are used, column 5, lines 12-18).

It would have been obvious to one of ordinary skill in the art at the time of invention to modify Hattori to extract more features from an input phoneme, thereby creating a higher dimensional pattern, and represent the space as a hypersphere, so that each pattern would only need two characteristic values to represent it, as taught by Kuhn (center point M and radius D, column 5, lines 45-51).

In regard to claims 26 and 32, Hattori discloses comparing a first distance to a first point associated from the received phoneme with a second distance associated in turn with each of the stored phonemes (the distance between an received vector and each reference vector is determined, and the distances are compared to find the minimal distance, which is used as the recognition result, column 3, column 3, lines 2-8 and lines 20-42).

Hattori does not disclose comparing distance from the center of the n-dimensional space.

Kuhn et al. disclose a method for comparing distances between received speech patterns in n-dimensional space (where  $n > 2$ , column 5, lines 8-18), comprising comparing the distance of a speech sample to the center point of a reference pattern, (column 15, lines 8-18).

It would have been obvious to one of ordinary skill in the art at the time of invention to modify Hattori to extract more features from an input phoneme, thereby creating a higher dimensional pattern, and to use a distance measurement to the center of the n-dimensional space, since the center point provides an "average" of the

Art Unit: 2655

extracted features to make a comparison to. This would ensure that the recognition of a phoneme would be dependent on the overall pattern of an input phoneme, rather than being adversely affected by a small number of divergent features.

In regard to claim 27, Hattori does not disclose "n" is at least approximately 100.

Kuhn et al. disclose that, in practice, higher dimensional characteristics are derived from speech samples so that a characteristic volume of the speech samples can be created (column 5, lines 14-18).

It would have been obvious to one of ordinary skill in the art at the time of invention to modify Hattori to make "n" at least approximately 100 dimensions, in order to provide a highly accurate characteristic volume for input phoneme samples, thereby increasing recognition accuracy.

In regard to claims 28, 29, 33, and 34, Hattori discloses:

determining a difference between a first distance and second distance for each stored phoneme; and

recognizing the received phoneme according to the stored phoneme associated with the smallest distance between the first distance and second distance.

The distance between a received vector and each reference vector is determined, and the distances are compared to find the minimal distance, which is used as the recognition result (column 3, column 3, lines 2-8 and lines 20-42). The comparison includes a differential vector calculation (column 4, lines 18-23).

10. Claim 17 is rejected under 35 U.S.C. 103(a) as being unpatentable over Hattori, in view of Kuhn et al., and further in view of Lilly et al.

Neither Hattori nor Kuhn et al. disclose using singular value decomposition.

Lilly et al. disclose using singular value decomposition (SVD) on extracted features from speech reduces the noise levels in the speech signal (which is common in real world speech recognition applications) and improves the recognition performance (page 259-260, section 4.2).

It would have been obvious to one of ordinary skill in the art at the time of invention to further modify the combination of Hattori and Kuhn et al. to use singular value decomposition, since using singular value decomposition in a front-end preprocessor (feature extractor) before recognition improves the recognition performance (page 260, section 5, lines 1-4).

11. Claims 23 and 24 are rejected under 35 U.S.C. 103(a) as being unpatentable over Hattori, in view of Kuhn et al., and further in view of Casey.

Hattori discloses converting the stored phoneme vector and the sample phoneme vector into n-dimensional space (see Fig. 3, input phonemes 14 and 15 and stored phonemes A, B, and C are converted to feature space).

Hattori does not disclose the n-dimensional space has a center and the probability density of the stored phonemes is approximately spherical.

Kuhn et al. disclose an n-dimensional space that has a center and the probability density of the stored phonemes is approximately spherical (points on a characteristic surface are situated at an approximate radius D from center point M, column 5, lines 42-45).

It would have been obvious to one of ordinary skill in the art at the time of invention to modify Hattori to extract more features from an input phoneme, thereby creating a higher dimensional pattern, and to use a distance measurement to the center of the n-dimensional space, since the center point provides an "average" of the extracted features to make a comparison to. This would ensure that the recognition of a phoneme would be dependent on the overall pattern of an input phoneme, rather than being adversely affected by a small number of divergent features.

Neither Hattori nor Kuhn et al. disclose the stored phoneme vector or the sampled phoneme vector have approximately 160 features.

Casey discloses a method for extracting features from a signal that is useful for phoneme recognition (column 4, lines 25-29). The method produces 40 bandpass signals windowed at every 20 msec. Given that the average length of a phoneme is approximately 125 msec, this would translate to approximately 160 parameter phoneme vectors (40 filters\*5 windows = 200 features).

It would have been obvious to one of ordinary skill in the art at the time of invention to further modify the combination of Hattori and Kuhn et al. to create stored and sampled phoneme vectors with approximately 160 parameters, in order to accurately represent both the time and frequency components of the phonemes.



**Conclusion**

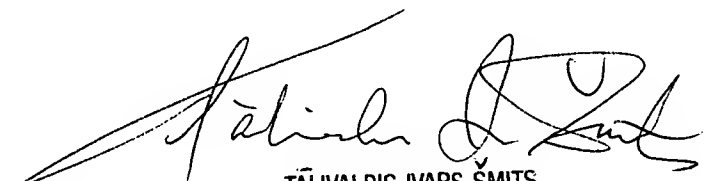
12. The prior art made of record and not relied upon is considered pertinent to applicant's disclosure. Watari et al. (U.S. Patent 4,601,054) disclose an additional method of comparing the distance between an input speech pattern and a stored speech pattern. Suzuki et al. (U.S. Patent 4,078,154) disclose a system for speaker identification that uses the center (centroid) of an input speech pattern.

13. Any inquiry concerning this communication or earlier communications from the examiner should be directed to Brian L Albertalli whose telephone number is (571) 272-7616. The examiner can normally be reached on Mon - Fri, 8:00 AM - 5:30 PM, every second Fri off.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Talivaldis Smits can be reached on (571) 272-7628. The fax phone number for the organization where this application or proceeding is assigned is 703-872-9306.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see <http://pair-direct.uspto.gov>. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free).

BLA 4/4/05



TĀLIVALDIS IVARS SMITS  
PRIMARY EXAMINER